Nigerian Journal of Scientific Research (NJSR)
ISSN-0794-0378

ON THE CHOICE OF THE PREDICTIVE PERFORMANCE OF PENALIZED RIDGE, LASSO AND ELASTIC NET ESTIMATORS USING SIMULATED DATA

USMAN, M.[1]*, SANI, S.S.[2], SANI, M.N.[1], BUKAR, S.[3], HAMZA, M.[3] AND ADAMU, K.[4]
[1]Department of Statistics, [2]Department of Agronomy, [3]Division of Agricultural College, Ahmadu Bello University, Zaria, Nigeria, [4]Department of Statistics, Nuhu Bamalli Polytechnic, Zaria, Nigeria

ABSTRACT

Penalized regression methods have been developed to overcome the drawbacks of ordinary least squares (OLS) regression model of its inability to perform well with respect to both prediction accuracy and model complexity, especially when the data suffer from multicollinearity problem and large sample size. Penalized model performs variable selection to reduce the number of parameters in the model in order to increase interpretability and reduce the model complexity. This study investigates the predictive performance of Ridge, LASSO and Elastic net penalized (or regularized) regression models. Monte Carlo simulation study was performed to obtain three sample sizes of the data (25, 50 and 100) with three predictors (12, 20 and 25). Root Mean Square Error (RMSE) was used to evaluate the performance of the three estimators. LASSO estimator outperformed Ridge and Elastic net in terms of prediction accuracy, since it exhibits a least RMSE from the overall results and can be recommended to practitioners.

Keywords: Penalized Regression, Multicollinearity, MSE, MLE, Simulation Study
***Correspondence:** ummohammed67@yahoo.com, 08035940877

INTRODUCTION

Advancement in technology makes it feasible to have access on large and complex datasets in different field of human endeavors. This huge volume of data is used by many organizations to arrive at a decision or to improve the organizational procedures and policies. The statistical estimation methods, under the umbrella of penalized methods, such as Ridge, least absolute shrinkage and selection operator (LASSO) and Elastic net that are well adapted to high dimensional problems are applied to such datasets. High dimensional data occur in a situation where the number of predictor variables is larger than the number of observations and this type of data occur in many fields with increasing dimensions being generated in those fields, such as medicine, engineering, agriculture, genetics, social sciences, signals and spectra in chemical analysis, image data in computer vision or microarray in gene data etc. According to [1] high-dimensional data have posed new challenges to statistical analysis, because modeling introduces overfitting, estimation instability and computational difficulty. These so called ill-conditioned problems, cannot be solved by OLS or Maximum Likelihood Estimation (MLE) methods. Scientists in different fields such as Biology, Finance, Statistics, social media, etc., collect high-dimensional data, which describes datasets with a high number of features or predictors exceeding the sample size [2]. In high dimensional data, predictor variables can be highly correlated.

When variables are highly related to one another, they do not provide unique or independent information to the regression analysis and for a reliable regression analysis, it is important to have predictor variables that are as unrelated to each other as possible. If some variables are so similar to each other, the regression can have difficulties to decide on the variables that are responsible for change in the response variable. With high dimensional data, penalized estimators that are often called shrinkage estimators, are used to obtain estimates in regression analysis. In this scenario the usual loss function is augmented with some type of penalty function that constrains or shrinks the coefficients in the model [3]. When predictor variables are highly related to one another or are highly dimensional, applying MLE estimation method results in large variance. Linear regression generally, is often used to build a model for predicting future outcomes or to find a link between the outcome and the predictor variables.

The accuracy of model's prediction and its complexity is critical in regression problem. The least squares regression is known for undesirable outcomes in terms of model complexity and prediction accuracy. Ridge regression [4], LASSO [5], Elastic net [6], (Bridge regression, [7] are among the regularised regression approaches developed to solve the faults of least squares regression [8]. Penalized regression method is applied for high dimensional data to simultaneously perform variable selection and estimate regression coefficients simultaneously. A well-known method for combating the challenge of

high-dimensionality curse is to solve a penalized regression problem, which combines the Residual Sum of Squares (RSS) loss function that measures the goodness of the fitted model to the datasets with some penalization terms that promote the underlying structure. Therefore, penalized methods can analyse and provide a good fit for the high dimensional data [9]. A typical example of the problem of multicollinearity can be observed in the manufacturing of cement where the amount of different compound in a hard brick is regressed on the heat developed from the cement [10]. Regularization penalty-based models, which include Ridge regression, Lasso and Elastic net, solve the problem of multicollinearity and high dimensionality. Despite not being new, these techniques are relatively less used by statisticians who traditionally tend to adopt more classic approaches [11].

MATERIALS AND METHODS

The penalized methods constrain or regularize the regression coefficients by shrinking them toward zero or exactly zero in the model. They incorporate penalty in the loss function to decrease the influence of multicollinearity variables in the model thus, improving the performance of the regression in the presence of variables that are highly related. Decrease in the estimate's variance can be achieved by introducing some amount of bias to shrink the regression coefficients towards zero and penalized regression methods are also, considered as shrinkage estimators. This study explores three penalized regression estimators – Ridge, LASSO and Elastic net.

Ridge regression

Ridge regression is one of the popular shrinkage estimators. The estimator was introduced by [12] and popularized by [4]. The aim of this method is to prevent multicollinearity problems in the regression. Ridge regression is also, known as Tikhonov regularization method, it provides a remedy for an ill-conditioned of the matrix form $X^T X$.

If $X = (x_1, x_2, ..., x_n)^T \in R^{nxp}$ is a matrix with column rank less than *p,* then least-squares regression will have problem [13]. In multiple regression analysis, the presence of multicollinearity which indicate the linear relationship among the independent variables can pose challenges in interpreting the results [14]. Given the general linear regression model:

$$Y = X\beta + \varepsilon, \qquad (1)$$

where $y = (y_1, y_2, ..., y_n)^T \in R^n$ is a vector of response variable, $X = (x_1, x_2, ..., x_n)^T \in R^{nxp}$ is a matrix of observations on the predictor variables, $\beta = (\beta_1, \beta_2, ..., \beta_n)^T$ is a vector of unknown regression coefficients and is a vector of error terms with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon^T) = \sigma^2 I_n$ where I_n is the unit matrix of order n and σ^2 is an unknown constant parameter. We assumed that the response variable is centred and the predictor is standardized, then:

$$\sum_{i=1}^{n} y_i = 0 \quad \sum_{i=1}^{n} x_{ij} = 0 \quad \sum_{i=1}^{n} x_{ij}^2 = n \qquad \text{for } j = 1, 2, \ldots, p$$

Therefore, Ridge regression imposes a constraint on the parameters and the parameters are estimated by minimizing the penalized sum of squares. Ridge regression shrinks the OLS estimators using the L2 norm of the coefficients. The Ridge estimate is given in matrix form as:

$$\hat{\beta}_{Ridge} = \arg\min_{\beta} \| y - x\beta \|_2^2 + \lambda \| \beta \|_2^2 \qquad (2)$$

$$\hat{\beta}_{Ridge} = \arg\min \left[\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \right] \qquad (3)$$

where *n* is the number of observations, $\lambda > 0$ is a regularized parameter and *p'*s are the number of predictors. When $\lambda = 0$, equation (3), becomes least squares regression.

K-fold cross-validation method is used to estimate the regularized parameter λ .

Now, for real valued function *f,* with domain *S*, arg min [$f(\beta)$] $\in$ $Sf(\beta)$ is the set of elements in *S* that achieve the global minimum in *S*. Equation (2) can also, be expressed as:

$$Q(L) = \sum_{i=1}^{n} (y_i - x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \qquad (4)$$

The solution can be obtained as:

$$\hat{\beta}_{(ridge)} = (X^T X + \lambda I)^{-1} X^T y \qquad (5)$$

where ***I*** is the *p x p* identity matrix. Adding positive value of λ being a positive scalar parameter, to the diagonal of X^TX makes the problem non-singular matrix making it invertible even with the multicollinearity in the data. The intercept β_0 is usually not included in the penalty function. This can be done by first centering the inputs and response variables. To select the regularized parameter λ in order to compute equation (5), a sequence of λ values (λ = 0.01, 0.02, 0.03, ….,1) are assigned and choose the optimal λ that minimises the mean squared error MSE(λ). The optimized λ is the one that produces the model minimizing the expected out-of-sample prediction error, often estimated by cross-validation method [15].

$$MSE(\lambda) = (Y - \widehat{Y}_{(\lambda)})^T (Y - \widehat{Y}_{(\lambda)}) / \mathrm{n} \quad (10)$$

where

$$\widehat{Y}_{(\lambda)} = X \widehat{\beta}_{(ridge)} \quad (7)$$

The optimal Ridge regularized parameter estimate $optimal\widehat{\beta}_{(Ridge)}$ is then given as,

$$optimal\hat{\beta}_{(ridge)} = (X^T X + \lambda_{opt} I)^{-1} X^T y \quad (8)$$

where λ_{opt} is the value of λ in which MSE(λ) attains the global minimum. In other words, λ_{opt} is the one with minimum MSE. The variance of the estimate is given by

$$Var(\hat{\beta}_{ridge}) = \sigma^2 diag[(X^T X + \lambda I)^{-1} X^T X (X^T X + \lambda I)^{-1}] \quad (9)$$

In contrast to the least square estimates, the Ridge estimator is biased, because the method introduces a bias to reduce the variance and the mean squared error (MSE) of the estimates to improve the prediction accuracy. Ridge estimator produces more stable estimates of the regression coefficients by shrinking the coefficients toward zero, but the method does not shrink the coefficients to exactly zero and therefore, the procedure does not perform variable selection to give an easily interpretable model.

Least absolute shrinkage and selection operator (LASSO) regression

LASSO was introduced by [5], being another penalized regression method that performs both variable selection and estimation of parameters in order to enhance the prediction accuracy and interpretability of the model. The rest of the variables which are not active factors can be removed from the model. LASSO improves the prediction accuracy and interpretation of model by altering the model fitting to select only a subset of predictor variables for use in the final model rather than using all of them. The research conducted by [16] compared Ridge, LASSO and Elastic net methods with various samples sizes using some levels of correlation. LASSO was found to have performed better. Lasso method constrains the sum of the absolute values of the regression parameters to be less than some value. The intercept β_0 is not included in the penalty. The inputs and response variables are centred first, and the model is fitted without the intercept. The Lasso penalty function can also be expressed in matrix form as given in equation (10):

$$\hat{\beta}_{LASSO} = \arg\min_{\beta} \| y - x\beta \|_2^2 + \lambda \| \beta \|_2 \quad (10)$$

$$\hat{\beta}_{LASSO} = \arg\min \left[\sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} \beta_j X_{ij} \right)^2 + \lambda \sum_{j=1}^{p} | \beta_j | \right] \quad (11)$$

where n is the number of observations, p is the number of predictors. λ is the regularized parameter estimated by cross-validation method. To obtain the $\boldsymbol{\beta}_{lasso}$ in matrix form, we need to minimise equation (11) with respect to $\boldsymbol{\beta}$. This involves differentiating the equation with respect to $\boldsymbol{\beta}$ and setting the derivative to zero in order to obtain the system of equation:

$$(X^T X)\beta + 0.5\lambda\beta^* = X^T Y \quad (12)$$

where $\boldsymbol{\beta}^*$ is defined as,

$$\boldsymbol{\beta}^* = \begin{pmatrix} \frac{\beta_0}{|\beta_0|} \\ \frac{\beta_1}{|\beta_1|} \\ \vdots \\ \frac{\beta_p}{|\beta_p|} \end{pmatrix} \quad (13)$$

It can be seen that equation (12) is not in closed form, so we carry out iterative method to determine the LASSO estimate of $\boldsymbol{\beta}$. Using the Newton Raphson Algorithm, and for a given $\lambda > 0$ and initial value of β_0 we run the iteration in equation (14):

$$\boldsymbol{\beta_t} = \boldsymbol{\beta}_{t-1} - (X^T X + 0.5\lambda \boldsymbol{G}_{t-1})^{-1} (X^T X \boldsymbol{\beta}_{t-1} + 0.5\lambda \boldsymbol{\beta}^*_{t-1} - X^T Y) \quad (14)$$

where the p +1 by p +1 diagonal matrix $\mathbf{G}$ is defined as,

$$\boldsymbol{G} = diag\left(\left(\frac{1-\beta_0^{-2}}{|\beta_0|}\right), \left(\frac{1-\beta_1^{-2}}{|\beta_1|}\right), \left(\frac{1-\beta_2^{-2}}{|\beta_2|}\right), \ldots\ldots, \left(\frac{1-\beta_p^{-2}}{|\beta_p|}\right) \right) \quad (15)$$

and t =1, 2, 3,… until convergence is achieved. A possible convergence criterion could be to stop the iteration whenever:

$$(\boldsymbol{\beta_t} - \boldsymbol{\beta}_{t-1})^T (\boldsymbol{\beta_t} - \boldsymbol{\beta}_{t-1}) / (\boldsymbol{\beta}_{t-1})^T (\boldsymbol{\beta}_{t-1}) < 10^{-6} \quad (16)$$

and take $\widehat{\boldsymbol{\beta}}_{LASSO} = \boldsymbol{\beta}_{t-1}$ for the given λ.

To obtain the optimal $\widehat{\boldsymbol{\beta}}_{LASSO}$, run equation (14) for a range of λ values (λ=0.01, 0.02, 0.03, ….,1) and choose the optimal $\hat{\beta}_{(lasso)}$ that minimizes the Mean Squared Error MSE(λ), while the corresponding value of λ gives the λ_{opt}

$$MSE(\lambda) = (Y - \widehat{Y}_{(\lambda)})^T (Y - \widehat{Y}_{(\lambda)}) / \mathrm{n} \quad (17)$$

and

$$\widehat{Y}_{(\lambda)} = X \widehat{\beta}_{(LASSO)} \quad (18)$$

where λ_{opt} is the value of λ in which MSE(λ) attains the global minimum. Alternatively, re-run the iteration

in equation (14) with λ_{opt} to obtain the optimal $\widehat{\boldsymbol{\beta}}_{lasso}$. The variance of $\widehat{\boldsymbol{\beta}}_{lasso}$ is given by:

$$Var(\hat{\beta}_{lasso}) = \sigma^2 diag[(X^TX + 0.5\lambda G)^{-1}X^TX(X^TX + 0.5\lambda G)^{-1}] \quad (19)$$

where

$$\sigma^2 = (Y - \widehat{Y}_{(\lambda)})^T (Y - \widehat{Y}_{(\lambda)}) / (n-p) \quad (20)$$

and n is the number of samples and p is the number of predictor variables.

The Shrinkage methods are now compared by examining the geometry of LASSO and Ridge regressions. The estimation problem for both methods is shown in Figure 1, with two predictor variables. The circular centered outlines around the OLS estimate show locations where the RSS is constant. The location where the elliptical contour intersects the constraint region is found by both Ridge and LASSO methods. Lasso has a blue diamond-shaped constraint region given by $|\beta_1| + |\beta_2| \leq t$ whereas Ridge regression has a green circle-shaped constraint region defined by $\beta_1^2 + \beta_2^2 \leq c$. In the case of LASSO, one of the coefficients β_j is equal to zero if the contour crosses the blue diamond at a corner [17].

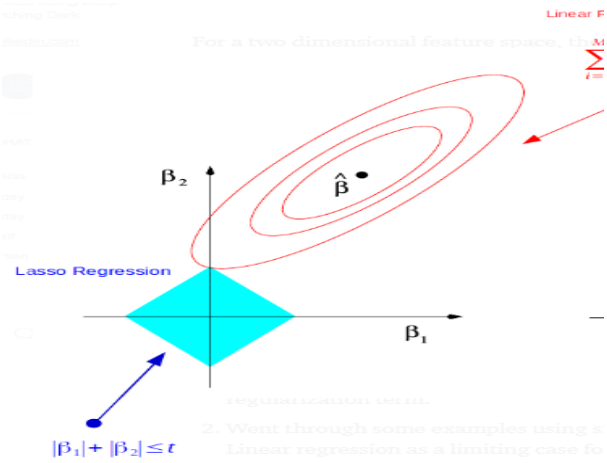

Figure 1: Lasso and Ridge geometrical interpretation: elliptical contours as the contours of errors and constraints for LASSO $|\beta_1| + |\beta_2| \leq t$ and Ridge $\beta_1^2 + \beta_2^2 \leq c$ [17]

Elastic net regression

Elastic net is another regularized and variable selection method which is the hybrid of Ridge and LASSO regression methods. It is used for the improvement of the selection, when groups of predictors are highly correlated. Elastic net procedure combines LASSO and Ridge regression by learning from their drawbacks to improve the regularization of the models [18], stated that, in estimating the parameter, the elastic net estimator found the Ridge regression coefficient and made the LASSO shrinkage along with the LASSO coefficient solution. Elastic net developed by [6] is as a result of critique of Lasso, whose performance of variable selection, depends much on data and makes it unstable. In elastic net regression, the estimated coefficients β are determined by minimizing a penalized sum of squares. In particular, the sum of squares in Elastic net is penalized by its penalty method which comprises of L1 and L2 norms, and is dependent on regularized parameters α and λ. Here, α is the weight that determines how much should be given to LASSO or Ridge, and λ alters the penalization weight [19]. The Elastic net can be obtained by the optimization problem:

$$\hat{\beta}_{Elastic} = \underset{\beta}{argmin}\|y - x\beta\|_2^2 + \lambda_1 \|\beta\|_1 + \lambda_2 \|\beta\|_2^2 \quad (21)$$

$$\hat{\beta}_{Elastic} = argmin\left[\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{p}\beta_j X_{ij}\right)^2 + \lambda\left[(1-\alpha)\sum_{j=1}^{p}\beta_j{}^2 + \alpha\sum_{j=1}^{p}|\beta_j|\right]\right] \quad (22)$$

where

λ is the tuning parameter

α is the weight that determines how much should be given to LASSO or Ridge regression, such that $0 \leq \alpha \leq 1$, where $\alpha = 0$, $\hat{\beta}_{Elastic\ net}$ becomes $\hat{\beta}_{Ridge}$ and where $\alpha = 1$, $\hat{\beta}_{Elastic\ net}$ becomes $\hat{\beta}_{Lasso}$

Equation (21) does not have a closed form and its solution can only be obtained iteratively using:

$$\boldsymbol{\beta_t} = \boldsymbol{\beta}_{t-1} - (X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}(X^TX\boldsymbol{\beta}_{t-1} + \lambda\{\alpha\boldsymbol{\beta}_{t-1} + 0.5(1-\alpha)\boldsymbol{\beta}^*_{t-1}\} - X^TY) \quad (23)$$

where $t = 1, 2, \ldots$ and the diagonal matrix G, β and β^* are as defined in equation (23). For a given α such that $0 < \alpha < 1$ and a given initial value β_0, we run equation (23) for each value of λ (λ =0.01, 0.02,…,1) until convergence is achieved based on the convergence criterion defined in equation (16), and then compute the MSE(λ). The λ_{opt} is obtained as that value of λ that returns the minimum MSE (λ). Having obtained λ_{opt} it can be used to re-run equation (23) with values of α say (α = 0.01, 0.02,…,0.99) and for each α we estimate MSE(α). The optimal α is determined as that α that returns the smallest MSE (α). Both the optimal α and λ are then used in equation (23) to re-run the iteration until convergence. The converged $\widehat{\boldsymbol{\beta}}$ is

the elastic net estimate $\hat{\boldsymbol{\beta}}_{elastic}$ of β. The variance of the estimate is then computed as,

$Var(\hat{\beta}_{elastic}) = \sigma^2\, diag[(X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}X^TX(X^TX + \lambda\{\alpha I + 0.5(1-\alpha)\boldsymbol{G}_{t-1}\})^{-1}]$ (24)

where σ^2 is defined as

$$\sigma^2 = (Y - \hat{Y}_{(\lambda)})^T (Y - \hat{Y}_{(\lambda)}) / (n-p) \quad (25)$$

and

$$\hat{Y}_{(\lambda)} = X\hat{\beta}_{(Elastic)} \quad (26)$$

The Elastic net estimator selects variables like LASSO, and shrinks the coefficients of correlated predictors like Ridge. It is found that this estimator is an extension of the Lasso and is robust to extreme correlations among the predictors [20]. The Elastic net estimator can select more than n variables when p is much greater than n [6]. Figure 2, shows the outmost contour shape of the Ridge penalty while the diamond shaped curve is the contour of the LASSO penalty. The yellow solid curve is the contour plot of the Elastic net penalty [6]. In the Elastic net procedure, the coefficient of Ridge method will be found first, then LASSO procedure will take place on the Ridge coefficient to shrink the coefficient for all the variables in the dataset [21].

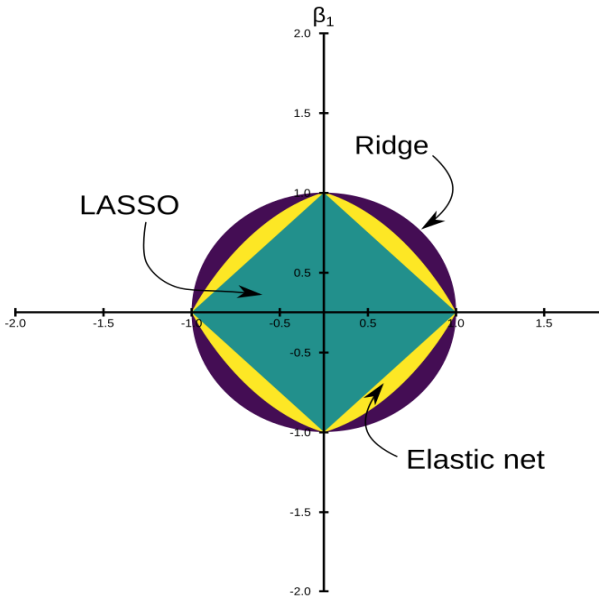

Figure 2: Geometry of the elastic net penalty

This study used simulated data to select the best estimator among the three estimators in terms of prediction accuracy.

The Monte-Carlo simulation

This study aims to investigate the predictive performance of three different penalized estimators (Ridge, Lasso and Elastic net) using simulated data. Root Means Square Error (RMSE) was used as a performance criterion. A Monte-Carlo simulation study has been conducted using R 2.4.3 programming language.

a) **Predictor variables**

Due to the fact that the degree of correlation among the predictor variables (X's) plays an important role in investigating the performance of the three estimators, the following four levels of correlation were used. ρ = 0.65, 0.75, 0.85 and 0.95

The n x p observations generated as:

U_{ij} for $i = 1, 2, \ldots, n$ $\quad j = 1, 2, \ldots, p$

$U_{ij} \sim N(0, 1)$

To generate the predictor variables (X's), the following equation was used:

$X_{ij} = (1 - \rho^2)^{1/2} U_{ij} + \rho\, U_{ip}$ (27)

$i = 1, 2, \ldots, n$ $\quad j = 1, 2, \ldots, p$

$X_{i1} = 1 \ \forall i$

Thereafter β_i is chosen for $i = 1, 2, \ldots, p$ randomly from $\beta_i \sim N(0, 1)$

b) Response variables

The response variables (y_i) is generated based on the following equation:

$$y_i = \exp(\beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \quad (28)$$

$$\log y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \quad (29)$$

Factors such as degree of correlation (ρ), sample size (n) and predictor variable (p) can affect the properties of the estimators. In this study, the objective is to investigate the predictive performance of the three estimators according to the strength of the multicollinearity, given the four different degrees of correlation between predictor variables. This study investigates the effect of sample size (n) on the predictive performance of the estimators. Sample size n = 12, 50 and 100 were considered and used. The number of predictor variables (X's) is important, for the fact that the negative impact of multicollinearity on the RMSE might be stronger when there are large number of predictor variables in the model. Also, the number of predictor variables p = 12, 20 and 25 were used in the study.

RESULTS AND DISCUSSION

Table 1, presents the Root Mean Square Error (RMSE) of the three estimators. It is clear, that values with red colour indicate that the Ridge model has the least RMSE. Values with blue colour indicate that LASSO model has the least RMSE. Values with green colour indicate that Elastic net has the least RMSE. From Table 1, we can see that, if the correlation is very high (ρ = 0.95) Elastic net perform better irrespective of the sample size of the data. When the number of predictor variables is high (p =25) with low correlation (ρ =0.65 and 0.75) Ridge model is better than LASSO and Elastic net, but when the correlation

is not so high (ρ =0.65 and 0.75), and the number of predictor variables (p=12 and 20) is also, not high, LASSO dominated the Ridge and Elastic net irrespective of the sample size. Table 1 clearly shows that the blue colour appeared more than any other colour.

Table 1: Results of the analysis with simulated data

CORRELATION	PREDICTOR	MODEL	RMSE		
			n = 25	**n = 50**	**n = 100**
ρ = 0.65	P = 12	RIDGE	13.7841	14.5082	14.4733
		LASSO	**3.4940**	**13.9393**	**14.4084**
		ELASITIC NET	12.8324	13.9771	22.5413
	P = 20	RIDGE	11.4142	17.3915	17.3783
		LASSO	**8.2817**	**13.4996**	**16.1223**
		ELASITIC NET	8.9991	17.4334	17.4223
	P = 25	RIDGE	**15.1169**	1314..8330	**1395.4203**
		LASSO	15.2569	1329.8037	1538.5460
		ELASITIC NET	15.2287	**1307.7511**	1411.1686
ρ = 0.75	P = 12	RIDGE	**13.0146**	9.2320	26.8535
		LASSO	16.9247	**7.5056**	**26.6089**
		ELASITIC NET	14.8217	7.5457	26.6918
	P = 20	RIDGE	**10.3292**	**12.5486**	42.8518
		LASSO	11.0774	12.6291	**41.4251**
		ELASITIC NET	10.4384	12.5486	41.5353
	P = 25	RIDGE	2432.128	**355.3153**	**3400.6714**
		LASSO	**77.38508**	361.3470	3430.3180
		ELASITIC NET	2472.5056	358.2919	3428.6272
ρ = 0.85	P = 12	RIDGE	41.2910	24.2890	16.1351
		LASSO	54.3229	24.2895	**15.7970**
		ELASITIC NET	**38.6369**	**15.6508**	15.8638
	P =20	RIDGE	**15.1164**	24.6039	12.2163
		LASSO	15.5578	**16.4710**	**11.4161**
		ELASITIC NET	15.1572	26.8569	13.2163
	P = 25	RIDGE	**83.8556**	17.0343	17,4417
		LASSO	84.2802	**15.3341**	17.4295
		ELASITIC NET	85.4388	17.6347	**16.4396**
ρ = 0.95	P = 12	RIDGE	7.0667	2.4068	1.8692
		LASSO	6.7102	2.4112	1.8703
		ELASITIC NET	**6.6992**	2.3982	**1.8685**
	P = 20	RIDGE	**3.3125**	2.6175	**8.2120**
		LASSO	4.3135	**1.4439**	8.2775
		ELASITIC NET	4.7185	1.5039	8.2598
	P = 25	RIDGE	**8.0624**	**4.8774**	1.8293
		LASSO	9.0034	5.2331	1.8191
		ELASITIC NET	9.0137	5.2481	**1.2293**

CONCLUSION

A Monte Carlo simulation study was performed to examine the performance of Ridge, LASSO and Elastic net regression models using root means square error (RMSE). A number of predictor variables were generated with different levels of correlation and sample sizes. Based on the results of the analysis with the simulated data, the findings have shown that, with high correlation (ρ = 0.85 and 0.95) and large sample size (n = 100), Elastic net use to perform better. In the case of small sample size (n = 12) and fairly strong correlation (ρ = 0.65 and 0.75), LASSO estimator performs better. When designing the experiment, factors such as the number of samples, the number of predictor variables and the degree of correlation have been changed. In the overall results of the analysis, the study found that LASSO estimator outperformed Ridge and Elastic net estimators, since it shows a least RMSE and can be recommended to practitioners.

REFERENCES

1. CHOOSAWAT C., REANGSEPHET O., SRISURADETCHAI P. & LISAWADI S. (2020). Performance Comparison of Penalized Regression Methods in Poisson

Regression under High-Dimensional Sparse Data with Multicollinearity. *Thailand Statistician,* **18**(3): 306-318.

2. HASNA, C., ASMAA, B. & TAYEB, O. (2024). Elastic net-based high dimensional data selection for regression. Expert Systems with Applications. *Elsevier,* **24**(4).
3. JOHAN DE, R. (2013). Penalized Estimation in High-Dimensional Data Analysis. PhD Thesis, Erasmus Universiteit Rotterdam op gezag van de rector magnificus: 1-4.
4. HOERL, A.E. & KENNARD, R.W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics,* **12**: 55–67.
5. TIBSHIRANI, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society.* Series B (methodological), **58**(1): 267-288
6. ZOU, H. & HASTIE, T. (2005). Regularization and variable selection via the elastic net. *Journal of Royal Statistical Society B*, **67**(2):301-320
7. FRANK, I.E. & FRIEDMAN, J.H. (1993). A statistical view of some chemometrics regression tools. *Technometrics*, **35**:109 – 148
8. PAREEKHAN A.O. (2022). Improving Prediction Accuracy of Lasso and Ridge Regression as an Alternative to Least Square Regression to Identify Variable Selection Problems. *ZANCO Journal of Pure and Applied Sciences*: 33-35
9. MAHDI, R., MONIREH, M. & NUR ANISAH, M. (2023). Penalized least squares optimization problem for high-dimensional data. *Int. J. Nonlinear Analysis and Application,* **14**(1): 245–250.
10. MUNIZ G. & KIBRIA B. M. G. (2009). On Some Ridge Regression Estimators: An Empirical Comparison *Communication in Statistics. - Simulations and Computation***, 38**(3): 621-630
11. YOKOCHI, C., BISPO, R. & RICARDO, F. (2023). Regularization Methods for High-Dimensional Data as a Tool for Seafood Traceability. *Journal of Statistic and Theory Practical,* **17(**44).
12. HOERL, A.E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, **58**(3): 54–59.
13. HASTIE, T. (2020). Ridge Regularization: An Essential Concept in Data Science. *Technometrics,* **62**(4): 426–433.
14. SAMIR, K.S., MOUZA, A., MAITHA, A. RUWAN, S. DANIA, S. & ZAINAB, N. (2023). Optimizing Linear Regression Models with Lasso and Ridge Regression: A Study on UAE Financial Behavior during COVID-19. *Migration Letters,* **20(**6): 139-153.
15. HANA S., GEORG H., ROK B. & ANGELIKA G. (2021). To tune or not to tune, a case study of ridge logistic regression in small or sparse datasets. *BMC Medical Research Methodology,* **21:** 199.
16. ALTELBANY, S. (2021). Evaluation of Ridge, Elastic Net and Lasso Regression Methods in the Precedence of Multicollinearity Problem: A Simulation Study. *Journal of Applied Economics and Business Studies,* **5**(1): 131–142.
17. EMMERT-STREIB, F. & DEHMER, M. (2019). High-dimensional LASSO-based computational regression models: Regularization, shrinkage, and selection. *Machine Learning and Knowledge Extraction,* **1**(1): 359-383.
18. ARAVEEPORN, A. (2021). The Higher-Order of Adaptive Lasso and Elastic Net Methods for Classification on High Dimensional Data. *Mathematics, **9***(10).
19. MOXLEY, T.A., JOHNSON-LEUNG, J, SEAMON, E., WILLIAMS, C. &, RIDENHOUR, B.J. (2023). Application of Elastic Net Regression for Modeling COVID-19 Sociodemographic Risk Factors.: 1-3.
20. FRIEDMAN, J., HASTIE, T. & TIBSHIRANI, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Sstatistical Software*, **33**: 1–20.
21. YUGESH, V. (2021). Hand-on- Tutorial on Elastic net Regression.